Capturing Location-Privacy Preferences: Quantifying Accuracy and User-Burden Tradeoffs

Michael Benisch **Patrick Gage Kelley** **Norman Sadeh**
Lorrie Faith Cranor

March 2010
CMU-ISR-10-105

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

We present a three-week user study in which we tracked the locations of 27 subjects and asked them to rate when, where, and with whom they would have been comfortable sharing their locations. The results of analysis conducted on over 7,500 hours of data suggest that the user population represented by our subjects has rich location-privacy preferences, with a number of critical dimensions, including time of day, day of week, and location. We describe a methodology for quantifying the effects, in terms of accuracy and amount of information shared, of privacy-setting types with differing levels of complexity (e.g., setting types that allow users to specify location- and/or time-based rules). Using the detailed preferences we collected, we identify the best possible policy (or collection of rules granting access to one's location) for each subject and privacy-setting type. We measure the accuracy with which the resulting policies are able to capture our subjects' privacy preferences under a variety of assumptions about the sensitivity of the information and user-burden tolerance. One practical implication of our results is that today's location sharing applications may have failed to gain much traction due to their limited privacy settings, as they appear to be ineffective at capturing the preferences revealed by our study.

This work has been supported by NSF grants CNS-0627513, CNS-0905562, and DGE-0903659. Additional support has been provided by Nokia, France Telecom, Google, the CMU/Microsoft Center for Computational Thinking, ARO research grant DAAD19-02-1-0389 to Carnegie Mellon University's CyLab, and the CMU/Portugal Information and Communication Technologies Institute.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAR 2010		**00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Capturing Location-Privacy Preferences: Quantifying Accuracy and User-Burden Tradeoffs	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Carnegie Mellon University,School of Computer Science,Pittsburgh,PA,15213**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
see report

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**22**	

Keywords: Expressiveness, Usable privacy, Location sharing, Web services, Social networking, Mechanism design

1 Introduction

The past few years have seen an explosion in the range of websites allowing individuals to exchange personal information and content that they have created. These sites include location-sharing services, which are the focus of this paper, social-networking services, and photo- and video-sharing services. While there is clearly a demand for users to share this information with each other, there is also substantial demand for greater control over the conditions under which their information is shared. This has led to expanded privacy and security controls on some services, such as Facebook, but designers of others appear reluctant to make this change. One reason for this reluctance may be that more complex privacy settings typically lead to more complex and hard-to-use interfaces.

Around one hundred different location-sharing applications exist today. These applications allow users to share their location (frequently, their exact location on a map) and other types of information, but have extremely limited privacy settings. Typically, they only allow users to specify a *white list*, or a list of individuals with whom they would be willing to share their locations at any time [21]. Despite the number of these types of applications available, there does not seem to be any service that has seen widespread usage. One possible explanation for this slow adoption has been established by a number of recent papers, which demonstrate that individuals are concerned about privacy in this domain [5, 7, 8, 13, 14, 18, 22]. However, our work is the first, to our knowledge, to study location-privacy preferences at a detailed enough level to address the question of whether or not more complex privacy-setting types may help alleviate these concerns.

We present the results from a user study where we tracked the locations of 27 subjects over three weeks in order to collect their stated location-privacy preferences in detail. Each day, for each of the locations a subject visited, we asked whether or not he or she would have been willing to share that location with each of four different groups: close friends and family, Facebook friends, the university community, and advertisers. Throughout the study, we collected more than 7,500 hours of location information and corresponding privacy preferences. In contrast to some earlier research that identified the requester's identity [7] and user's activity [6] as primarily defining privacy preferences for location sharing, we find that there are a number of other critical dimensions in these preferences, including time of day, day of week, and exact location.

We characterize the complexity of our subjects' preferences by measuring the accuracy of different privacy-setting types. We consider setting types that allow a user to share his or her location based on the group of the requester, the time of day of the request, whether or not the request is made on a weekend, and his or her location at the time of the request. Using the detailed preferences we collected during the location tracking phase, we identify each subject's most accurate collection of rules, or *policy*, under each privacy-setting type. To test the effectiveness of the different setting types, we measure the accuracy with which each is able to capture our subjects' preferences, while varying assumptions about the relative cost of revealing a private location, and about our subjects' tolerance for user burden.

As one might expect, we find that more complex privacy-setting types, such as those that allow users to specify both location- and time-based rules, are more accurate at capturing the preferences of our subjects under a wide variety of assumptions. More surprising is the magnitude of accuracy improvement — in some cases more complex setting types can result in almost three times the average accuracy of white lists. White lists appear to be particularly ineffective at capturing our subjects' preferences. Even relatively simple extensions, such as those that allow rules based only on time of day, can yield a 33% increase in average accuracy, if we assume that our subjects are privacy sensitive. This finding is also consistent with results from our pre-study survey, where subjects reported being significantly more comfortable with the prospect of sharing their location using time- and location-based rules, compared to white lists.

In addition to accuracy, we measure the amount of time each day that our subjects would have shared their location under each of the different privacy-setting types. Interestingly, we find that more accurate setting types also

lead to more sharing. This result, which at first may seem counter intuitive, actually makes sense: when users have complex privacy preferences and are given limited settings, they generally tend to err on the safe side, which causes them to share less. This may explain why some social networking sites, such as Facebook, have begun to move toward more complex privacy-setting types — if users end up sharing more, the services are more valuable. The lack of sharing we observe with simple setting types may also help explain the slow adoption of today's location sharing applications.

While our results suggest that more complex privacy-setting types are necessary to capture the true location-privacy preferences of the user population represented by our subjects, these settings do not come without cost. More complex setting types generally imply additional user burden, especially if they require users to specify significantly more rules than their simple counterparts. To address this, we examine a number of different privacy-setting types, which range from being fairly simple to more complex, under varied assumptions regarding the amount of effort our subjects would be willing to exert while creating their policies. For the purposes of this paper, we use the number of rules a policy contains as a proxy for the user burden involved in specifying it. Our findings suggest that, while limiting policies to a small number of rules dampens the accuracy benefits of complex privacy-setting types, they generally remain substantially more accurate than white lists.

The user study presented in this paper can be generalized as a methodology for characterizing the tradeoffs between more complex setting types and accuracy in a number of privacy and security domains. At a high level, the methodology involves i) collecting highly detailed preferences from a particular user population, ii) identifying policies for each subject under a variety of different privacy- or security-setting types, and iii) comparing the accuracy of the resulting policies under a variety of assumptions about the sensitivity of the information and tolerance for user burden.

The rest of this paper proceeds as follows. In the next section, we present an overview of related work on location sharing and privacy preferences. In Section 3, we provide the details of the methods used in conducting our user study and analyzing the data. In Section 4, we present a detailed analysis of our data. Finally, we present some conclusions and possibilities for future work in Section 5.

2 Related work

Location-sharing services are an area of significant growth as consumers gain access to ever cheaper and "smarter" mobile phones. With expanding market share, these services are anticipated to capture a significant portion of the billions of dollars in marketing revenue from the broader class of location-enabled applications [10]. Yet, despite analyst predictions and the growing number of location-sharing applications that have been developed, no service has captured a significant market share.

While high-profile services that are built around location sharing, like Loopt[1] and Google's Latitude,[2] seem to dominate the press, neither has been crowned a "killer app." Dozens of other offerings exist, many built around technology platforms that have allowed easier creation of these applications, including the iPhone SDK,[3] and Google's Android SDK,[4] as well as Yahoo's FireEagle Platform, which as of March 2010, has 79 applications in its gallery.[5] The FireEagle platform facilitates privacy-enhanced sharing by allowing users to specify a policy for each service that he or she provides with access. FireEagle allows, just as Google's Latitude, exact-location or city-level granularity sharing with white-listed entities. However, Tsai *et al.* found that privacy protection through the abstraction

[1] Loopt. http://loopt.com/
[2] Latitude. http://www.google.com/latitude
[3] iPhone Dev Center. http://developer.apple.com/iphone/
[4] Android. http://code.google.com/android/
[5] Fire Eagle. http://fireeagle.yahoo.net/

of location is rare. Of 89 sharing services surveyed in that work, only 11 provided any control over the granularity of the location disclosure, while over half of the services (50) used white-listing (or, equivalently, black-listing) to protect a user's location [21]. They also found that more complex privacy-setting types were nearly nonexistent in the landscape at the time, with only 11 services providing group designations, and only two having approvals with expirations. One notable exception was Locaccino,[6] which was developed by our research group at CMU and allows users to specify time- and location-based rules (these are richer privacy settings than those offered by any commercial service).

Many research groups have developed location-based services, including PARC's Active Badges [24], ActiveCampus [2], MyCampus [17], Intel's PlaceLab [11], and MIT's iFind [12]. However, the research done with these systems rarely reached the point of studying privacy preferences. Instead this work was typically hampered by adoption and technological issues. Work on a Semantic Web framework to capture rich privacy preferences in different context-aware applications, including location sharing applications, was also conducted in the context of CMU's MyCampus project [17]. This work later led to the development of several other location sharing applications at CMU, including PeopleFinder [18], and most recently Locaccino.

As far back as 2003, users of a diary study cited some concerns about location privacy, stating a preference to not have their phones tracked [2]. A study using the experience sampling method in 2005 found that location-privacy preferences were complex, and "participants want to disclose what they think would be useful to the requester or deny the request" [7]. These findings are evidence that without more complex privacy-setting types, users will simply shutdown, and deny requests if they cannot specify policies that would lead to useful sharing. One drawback of this research is that much of it focused on laboratory experiments [8, 16] and small group testing [1, 13, 20], where there are minimal privacy concerns given the small number of (often simulated) requests.

As far as we know, there have been only two other field studies that revealed complexity in people's location-privacy preferences. The first, by Tsai *et al.*, found that having feedback, or information on who had viewed one's location, had a significant impact on how comfortable people were with sharing their information [22]. Burghardt *et al.* went further by exposing individuals to five different privacy technologies in a real world deployment. They reported findings related to both subjects' preferences among the different technologies, and the effectiveness of the technologies [5]. The findings of these two studies are similar to ours, in that they suggest users have rich location-privacy preferences; however, they did not capture these preferences in as much detail as we do. For example, Burghardt *et al.* asked subjects, prior to being tracked, to report locations they considered private. They then treated these reports as the subjects' true privacy preferences. Not only is this method less detailed than ours, it is also problematic given Connelly *et al.*'s findings [6], which indicated that subjects tended to have significant differences between previously asserted and *in situ* privacy preferences.

The fact that more complex privacy and security settings are needed to capture people's preferences has been observed in other domains as well. For example, Mazurek *et al.* observed that people needed fine-grained access control for configuring their file-sharing preferences [15]. The benefits of more complex forms of expression have also been studied more generally in the field of economic mechanism design [3, 4].

3 Methods

In this section we provide an overview of our study, details of the software we used to conduct it, descriptions of the privacy-setting types we consider, and a description of the methods we use to analyze them.

[6] Locaccino. `http://locaccino.org/`

3.1 Study overview

The data for our study was collected over the course of three weeks in early November 2009. We supplied 27 participants with Nokia N95 cell phones[7] for the entire study. Each subject was required to transfer his or her SIM cards to the phone we provided and use it as a primary phone at all times. This requirement ensured that subjects kept the phones on their person, and charged, as much as possible. Each of the phones was equipped with our location-tracking program, which recorded the phone's location at all times using a combination of GPS and Wi-Fi-based positioning.

Each day, subjects were required to visit our web site where the locations recorded by their phones were filtered into distinct location observations. For each location a subject visited, we asked whether or not he or she would have been comfortable sharing that location with different groups of individuals and advertisers. While no location sharing to others actually occurred, we solicited the names of people from the different groups (other than advertisers) so that the questions the subjects answered were more meaningful.

We also administered surveys before and after the study to screen for participants, measure the level of concern about privacy that people had about sharing their location information, and collect relevant demographics. The screening process ensured subjects had, or were willing to purchase, a cellular data plan with a compatible provider.

Subjects were paid a total of $50-$60, corresponding to $30 for their successful participation in the study day, and $20-$30 to reimburse them for the data plan that was required by the location-tracking software.

3.2 Software

The primary materials we used in our experiment included location-tracking software written for the Nokia N95 phones and a web application that allowed subjects to audit their location information each day.

3.2.1 Location-tracking software

Our location-tracking software is written in C++ for Nokia's Symbian operating system. It runs continuously in the background, and starts automatically when the phone is turned on. During normal operation, the software is completely transparent – it does not require any input or interaction.

When designing our software, we faced two primary challenges: i) managing its energy consumption to ensure acceptable battery life during normal usage, and ii) determining the phone's location when indoors or out of view of a GPS signal.

To address these challenges, our software is broken down into two modules: a *positioning module* that tracks the phone's location using a combination of GPS and Wi-Fi-based positioning, and a *management module* that turns the positioning module on and off to save energy.

Positioning module. To estimate the position of the phone, our positioning module makes use of the Nokia N95's built in GPS, and Wi-Fi units. When activated, the positioning module registers itself to receive updates from the GPS unit at a regular interval (15 seconds). When the GPS unit is able to determine the phone's position, the positioning module records its latitude and longitude readings.

In our initial tests we found that the GPS signal was unreliable when the phone was indoors, and even when the phone was outdoors on cloudy days. For that reason, whenever the positioning module is active, it also records the MAC addresses and signal strengths of all nearby Wi-Fi access points at a regular interval (3 minutes). We are able to use this information to determine the physical address of the phone with a service called Skyhook Wireless.[8]

[7] These phones were generously provided by Nokia.

[8] Details about the Skyhook API are available at http://skyhookwireless.com/.

While the positioning module is active, it sends all location information to our server using the phone's cellular data connection in real time.

Management module. Our initial tests revealed that leaving the GPS unit on continuously resulted in an unacceptable battery life of 5-7 hours on average. The management module uses the N95's built in accelerometer to address the issue of energy consumption. It constantly monitors this low energy sensor, and only activates the positioning module when the accelerometer reports substantial motion. In practice we found that this improved the phone's battery life to 10-15 hours on average.[9]

3.2.2 Web application

Each day, subjects were required to visit our web site to audit the locations they visited that day. The locations were first filtered, then presented to the subjects to audit.

Location filtering. When a subject logs into our web application, it iterates through each of the GPS and Wi-Fi readings that have been recorded since the last time the user audited his or her locations. Each of these readings is either aggregated into a location observation, if the user stood still, or a path observation, if the user moved.[10] A new location observation is created when a subject has moved more than 250 meters from his or her last known location and remained stationary again for at least 15 minutes.

Audit administration. After a subject's locations have been filtered, our web application takes the subject through a series of pages that trace his or her new locations in chronological order. Each page displays a location on a map, inside a 250-meter ring, indicating the subject's estimated location during a particular time period. The times when the subject arrived and departed from the location are indicated next to the map. Each page also includes a link that allows subjects to report that an observation was completely inaccurate (inaccurate observations accounted for about 2% of the time, and are removed during our analysis). A screen shot of the user interface for this part of the web application is shown in Figure 1 (left).

Underneath the map, our web application presents four questions, each corresponding to a different group of individuals. The right side of Figure 1 shows an example screen shot of a question for the friends and family group. Each question asks whether or not the subject would have been comfortable sharing his or her location with the individuals in one of the groups. The groups we asked about in our study were: i) close friends and family, ii) Facebook friends, iii) anyone associated with our university, and iv) advertisers. Subjects are given the option of indicating that they would have shared their location during the entire time span indicated on the page, none of the time span, or part of the time span (when part of the time is chosen, a drop down menu appears allowing the subjects to specify which part of the time they would have allowed, as shown in Figure 1).

Questions about the friends and family and Facebook groups included a fourth option, allowing subjects to indicate that they would have been comfortable sharing their location with some of the individuals in the group, but not all of them. This option was chosen about 20% of the time for the Facebook friends group. However, 89% of the time this option was chosen, the subject also reported that he or she would have been comfortable sharing with either friends and family, or the university community. These subjects were most likely considering one, or both, of these two groups as subgroups of Facebook friends. This hypothesis is further supported by the fact that 82% of the

[9] For more details about this process, see the description of a similar technique used by Wang *et al.* for managing energy consumption while tracking users with mobile devices [23].

[10] Path observations between locations were also depicted on some pages. However, we do not address those observations in this paper since they accounted for less than 1% of the observed time.

subjects reported in the post-study survey that they did not feel there were any relevant groups missing from our list. For these reasons, we treat the "Yes, for some of these people" response as denying the entire group in our analysis.

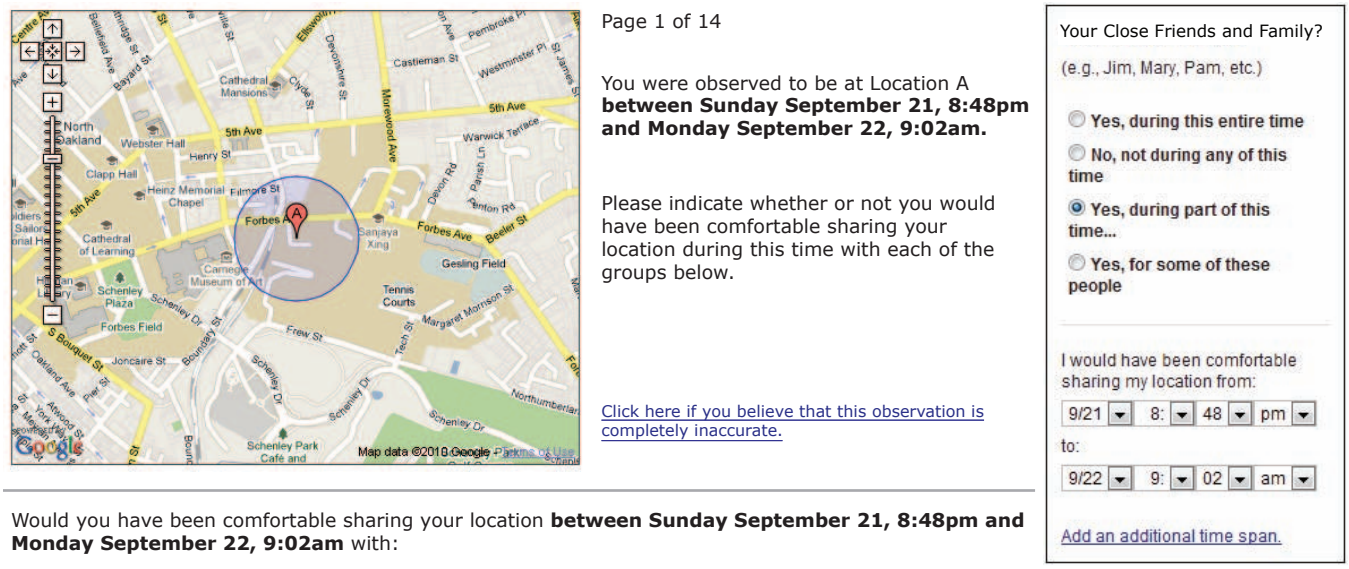

Figure 1: A screen shot of our web application displaying an example location on a map between 8:48pm and 9:02am (left), and an audit question asking whether or not a subject would have been comfortable sharing the location displayed on the map with the friends and family group (right). An audit question, like the one shown here, appeared below the map for each of the groups, at each location a subject visited.

3.3 Privacy-setting types we compare

In our analysis (Section 4.3), we focus on evaluating the accuracy of the following different privacy-setting types, which range from being fairly simple to more complex. We will illustrate the differences between these setting types by considering a hypothetical user named "Alice," who wishes to share her location only with her friends, when she is at home, and on the weekends, between the hours of 9am and 5pm. In the absence of a rule sharing one's location, we assume that the default behavior of a sharing service would be to deny.

- **White list.** White lists are the least complex privacy-setting type we consider. They only allows users to indicate whether or not they would be comfortable sharing their location with each group, at all times, and at all locations. The accuracy of white lists can be viewed as a measure of the importance of a requester's identity in capturing users' privacy preferences. White lists are user friendly, since they only require a single rule indicating who can view one's location.

  Using a white list, our hypothetical user, Alice, would need to indicate *who* (individually or by group) is allowed to see her location. Similarly, she may also create a rule that everyone is allowed to see her at all times with a list of exceptions (i.e., a black list). Alice's policy under this setting type would not match her preferences, since friends on her white list would be able to see her anytime and anywhere.

- **Location (Loc).** Loc settings allow users to indicate specific locations that they would be comfortable sharing with each group. Loc settings are more complex than white lists, since white listing a group can be simulated with Loc settings by sharing all locations with that group. The accuracy of Loc settings can be seen as a

measure of the importance of location in capturing users' privacy preferences. A single location rule is defined by a latitude-longitude (lat-lon) rectangle and a set of people or groups who can view the user's location within the rectangle.

Alice would need to create a rule allowing her friends to view her location when she is at home, by indicating it with a rectangle on a map. However, this policy does not match her preferences, since her friends could see that she is home at night or on a weekday.

- **Time.** Time settings allow users to indicate time intervals (discretized into half-hour blocks) during which they would be comfortable sharing their locations with each group (this setting type does not consider the day of the week). Similar to Loc settings, Time settings are more complex than white-lists, since white listing for an individual or group can be simulated by granting them access at all times. The accuracy of Time settings can be seen as a measure of the importance of the time of day in capturing users' privacy preferences. A single time rule is defined by a start time, an end time, and a set of people or groups who can view the user's location between the two times.

  With Time settings, Alice would need to create a rule sharing her location with her friends, between 9am and 5pm, regardless of where she was, and the day of week. Alternatively, she could err on the safe side and choose to share a smaller time window during which she feels she is more likely to be home. In either case, Alice's policy would not match her preferences, since her friends could see her location when she is somewhere other than at home.

- **Time with weekends (Time+).** Time+ settings are the same as Time settings, but they allow users to indicate time intervals that apply only to weekdays, only to weekends, or to both. The improvement in accuracy of Time+ over Time can be viewed as the importance of weekends in capturing our subjects' privacy preferences. A single rule with Time+ settings is defined by a start time, an end time, a flag indicating whether it applies to weekdays, weekends, or both, and a set of people or groups who can view the user's location, between the two times, on the specified type of day.

  With Time+ settings, Alice would need to create a rule sharing her location with her friends, between 9am and 5pm on weekends only, regardless of where she was. As with Time settings, Alice's policy would not match her preferences, since her friends could see her location when she is somewhere other than at home, but with Time+ settings this could not happen on a weekday.

- **Location and time (Loc/Time).** Loc/Time settings combine the Loc and Time setting types described above. They allow users to indicate time intervals during which they would be comfortable sharing specific locations with each group. The accuracy improvement of Loc/Time over Loc and Time individually can be viewed as the importance of offering both types of settings together. A single Loc/Time rule is defined by a start time, an end time, a lat-lon rectangle, and a set of people or groups who can view the user's location when he or she is within the rectangle, between the two times.

  With Loc/Time settings, Alice would need to create a rule allowing her friends to see her when she is at home, from 9am to 5pm, regardless of the day of week. In this case, Alice's policy would not match her preferences, since her friends can see her at home on a weekday.

- **Location and time with weekends (Loc/Time+).** Loc/Time+ settings are the same as Loc/Time settings, but they allow users to indicate time intervals that apply only to weekdays, only to weekends, or to both. This is the most complex privacy-setting type we consider.

  Using Loc/Time+ settings, Alice would be able to express her true privacy preferences with a single rule, allowing her friends to see her when she is at home, from 9am to 5pm, on weekends only.

3.4 Measuring accuracy with variable cost

In order to measure the accuracy of different privacy-setting types, we first identify a collection of rules, or a *policy*, for each subject, under each of the different types described in Section 3.3. For a subject, i, a privacy policy, p, and group, g, we define the accuracy of the policy for i and g using two functions, `correct_hrs` and `incorrect_hrs`. The functions take as input i, p, and g, and return the number of hours correctly shared and incorrectly shared, respectively, by subject i, with group g, under p. These statistics are easily computed from our data for any possible policy, since we can simulate what the policy would have done at each of the locations a subject visited, and compare that to their stated preferences for that location. We normalize the accuracy to be a fraction of the time shared by a policy that perfectly matches the subject's preferences (i.e., shares whenever the subject indicated he or she would do so, and does not share at any other times or locations), which we denote as p^*.

In our analysis, we will consider the accuracy of different privacy-setting types while varying assumptions about our subjects' tolerance for mistakes. For this, we define a penalty term, or cost, c, associated with mistakenly revealing a piece of private information. In our analysis, we vary c from 1 to 100, and investigate the impact it has on accuracy and sharing, under the different privacy-setting types. Varying c amounts to varying the ratio between the reward for revealing a location when a subject indicated that he or she would have shared it, and the penalty for revealing it when he or she indicated not being comfortable with having it shared. At the lowest level (when $c = 1$) these two occurrences are equally rewarded and penalized, respectively. When $c = 100$, mistakenly revealing a location is considered to be one-hundred times as bad as correctly revealing it. This level of cost is essentially equivalent to the assumption that our subjects would be very cautious, and never make policies that mistakenly revealed their locations. Varying this cost helps to account for differences between subjects and across potential applications.[11] Accuracy for a policy, group, and subject is given by the following equation.

$$\frac{\texttt{correct\_hrs}(i,p,G) - c \times \texttt{incorrect\_hrs}(i,p,G)}{\texttt{correct\_hrs}(i,p^*,G)} \tag{1}$$

The accuracy of the best policy for any subject, group, and privacy-setting type, will always be between zero and one. It can never be below zero, because an empty policy achieves zero accuracy, and it can never be above one, since we normalize the accuracy for each subject using the accuracy of the best possible policy for that subject.[12]

3.5 Identifying privacy policies with user-burden considerations

In Sections 4.3.1 and 4.3.2, we consider the most accurate policy for each subject, and given privacy-setting type, with no limit on the number of rules in the policy. We then consider the effect of limiting the number of rules to account for user-burden tolerance in Section 4.3.3. In both cases, the accuracy values that we report can be taken as upper bounds on the accuracy we would expect in practice, since subjects may not always create the most accurate possible policy.

With no rule limit, a subject's most accurate policy, for a given group and setting type, can be easily computed by identifying all possible *atomic rules* for that and setting type (e.g., rules that apply to a single location, or a single 30 minute block). We can then greedily add an atomic rule whenever it would result in positive accuracy for that subject (i.e., when it is correct more than $1/c$ of the time). This method is guaranteed to identify the most accurate

[11] We assume that there is no penalty for mistakenly withholding a location, since our post-study survey results suggest that subjects had relatively little dis-utility at this prospect. However, this can easily be added as an additional cost to the accuracy calculation in Equation 1.

[12] When a subject indicated that he or she would never have shared their location with a particular group, thereby making the accuracy equation undefined, we report the accuracy for that subject and group as one, since we assume that the default behavior of the system is to deny access, which is consistent with the subject's preferences.

policy, since the search for rules with no limit decomposes in a straightforward way: each group, time, location and location/time pair can be considered independently (when rules regarding weekends and weekdays are allowed, we treat times on the two types of days independently). For example, the effect on overall accuracy of adding a rule sharing a particular location does not depend on which other locations the policy ends up sharing.

Like many combinatorial problems (e.g., knapsack, job-shop scheduling, graph coloring), the problem of identifying the most accurate policy becomes substantially harder with a limited resource. With a limit on the number of rules, the greedy solution is no longer guaranteed to identify the most accurate policy. To address this problem, we developed a tree search technique, based on the well-known A* search algorithm, for computing a subject's most accurate policy with no more than k rules.

Each level of the search tree corresponds to one of the rules in the policy, and each branch represents a particular rule that can be included. For example, one branch could correspond to the rule "University community and Friends can see me at any location, between 8:00am and 7:00pm, on weekdays." Thus, at any node, j, with depth d, a policy with d rules can be constructed by traversing the edges from j to the root. Figure 2 illustrates part of a search tree using Loc/Time+ settings.

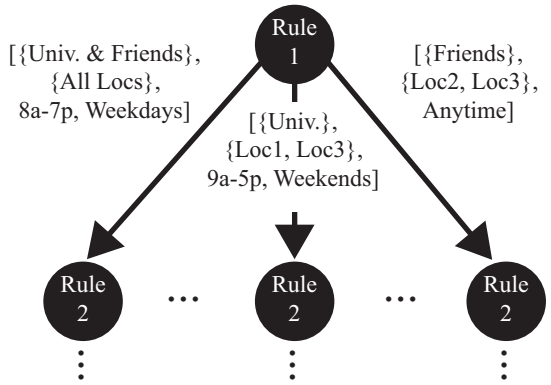

Figure 2: Part of a search tree for identifying a subject's most accurate privacy policy using Loc/Time+ settings.

Our search begins at the root node, and constructs one child node for each of the possible rules a user could add, given the type of settings available. The nodes are added to a priority queue, called the *open queue*. Nodes are then popped off the open queue one at a time until a leaf node (i.e., node with depth k) is reached. Whenever a node, j, is removed from the open queue, a child of j is added to the queue for each of the remaining *feasible* rules. A rule is considered feasible for inclusion in children of j if it does not *overlap* with any rule that is already in the policy represented by j. Two rules overlap if they refer to the same place, time, or place and time, for Loc, Time (Time+), and Loc/Time (Loc/Time+) settings respectively.

As usual, our search orders the nodes in its open queue according to an admissible (i.e., optimistic) heuristic. The heuristic approximates the accuracy of any policy with k rules originating from a particular node as the total accuracy of the rules included so far, plus the accuracy of a greedy solution over the remaining feasible rules with no rule limit. This approximation is guaranteed to be greater than or equal to the best total accuracy of any set of k rules that descends from node j. It may overestimate this value, if the greedy solution uses more than k rules. By using the A* node selection strategy, our search ensures that any node that it visits has a lower (or equal) accuracy than any previously visited node, thus making the first solution reached provably the most accurate possible.

If we were to consider every possible atomic rule at each level of this search tree it would be intractable for the more complex types of privacy settings. To address this, we losslessly compress the search space by preprocessing each subject's ground truth policy. For Loc rules, individual locations are grouped together into complex locations if they are audited the same way at all times (i.e., sharing them always results in positive accuracy for the same groups),

and it would be possible to draw a rectangle around them without including any of the subject's other locations. For Time (Time+) rules, individual 30-minute spans are grouped together if they are audited the same way every day (or type of day for Time+). For Loc/Time (Loc/Time+) rules, locations are grouped together if they are always audited the same way based on time of day, and it would be possible to draw a rectangle around them without including any other locations. With these preprocessing steps in place, we can identify policies for each subject, and setting type, typically in a matter of seconds.

4 Results

Before we present our analysis on measuring the effects of different privacy-setting types, we will describe our survey findings, the general mobility patterns we observed, and some high-level statistics that demonstrate the complexity of our subjects' locaiton-privacy preferences.

4.1 Survey results

Our 27 subjects were all students or staff at our university. The sample was composed of 73% males and 27% females, with an average age of about 22 years old. Undergraduates made up 58% of our sample, graduate students made up 35%, and two people (7%) were staff members.

In our pre-study survey, we asked participants about how comfortable they would be if close friends and immediate family, Facebook friends, members of the university community, or advertisers could view their locations at anytime, at times they had specified, or at locations they had specified. Based on ratings on a 7-point Likert scale (ranging from "not comfortable at all" to "fully comfortable"), we found that, in general, participants were more comfortable with their close friends and family locating them than their Facebook friends, people within their university community, or advertisers.

Within each group, we found that respondents had relatively equal levels of comfort for time-based or location-based rules (the differences were not statistically significant).[13] However, it is interesting to note that location had a substantially higher average score than time for the advertiser group, since we later find that this is the only group for which the difference between the accuracies of Loc settings and Time settings is marginally significant. The average scores for this question are shown in Table 1.

Group	Anytime	Location	Time
Friends and family	5.00	6.08	6.36
Facebook friends	3.64	4.88	5.40
University community	3.28	4.56	5.00
Advertisers	2.60	4.32	3.60

Table 1: The average report on our pre-study survey of how comfortable subjects would have been, on a 7-point Likert scale from "not comfortable at all" to "fully comfortable" if their location could be checked by each of the groups "Anytime," "At locations you have specified," or "At times you have specified."

We also found that subjects reported that they would be significantly more comfortable, on average, for the Facebook friends, university community, and advertiser groups, using location- and time-based rules than with white lists. For example, for the advertisers group, our subjects indicated that they would not be comfortable if their

[13] We use a two-sample independent t-test with unequal variances for all tests of statistical significance, unless otherwise noted. We report p values of less than 0.05 as significant, and less than 0.1 as marginally significant.

locations were shared all the time (M=2.6); but at times (M=3.60) or locations (M=4.32) they had specified, their comfort levels would significantly increase.

After completing our study, we asked our participants how bad they thought it would have been, on a 7-point Likert scale from "not bad at all" to "very, very bad," if the system had shared their information at times when they did not want it to be shared, or if the system had withheld their location when they wanted it to be shared. Table 2 shows the average report for each type of mistake and each group.

Group	Mistaken withhold	Mistaken reveal
Friends and family	3.00	3.26
Facebook friends	2.30	3.70
University community	2.07	4.26
Advertisers	1.67	4.74

Table 2: The average report of how bad subjects thought it would have been, on a 7-point Likert scale from "not bad at all" to "very, very bad," if their location were mistakenly withheld from or revealed to each of the groups.

Our subjects reported significant levels of dis-utility at the prospect of their locations being mistakenly shared with the university community, Facebook friends, and advertisers groups, with the worst being advertisers, where 33% of the participants chose 7 on the scale, and 50% choose 5 or more. In contrast, our subjects reported relatively little dis-utility at the prospect of their locations being mistakenly withheld. We also see an inverse relationship between the average report within groups, such that groups where mistakenly revealing is worse tend to have lower reports for mistakenly withholding. This lends support to the hypothesis that our subjects would tend to share less when given simpler privacy-setting types, since they report being far more concerned with inadvertent disclosure of their location than with it being withheld, on average.

We also asked our subjects how often they would have answered the questions differently if we had actually been sharing their locations. The majority of subjects (about 70%) responded that they would have rarely or never answered differently. Another 15% said they would have answered differently some of the time, and the rest said most or all of the time.

4.2 Mobility patterns and preference statistics

On average, our subjects were observed for just over 60% of the time during our experiment, and our observations were distributed relatively evenly throughout the day. We found that, on average, subjects would have been comfortable sharing their locations about 93% of the time with friends and family, 60% of the time with Facebook friends, 57% of the time with university community, and 36% of the time with advertisers.

Figure 3 shows how our subjects' preferences varied with time of day, and day of week. It shows the average percentage of time subjects were willing to share during each 30-minute interval, separately for weekdays and weekends. Preferences for the friends and family group are largely unaffected by time of day or day of week. However, the results show substantial variation in preferences based on time of day and day of week, for the other three groups. For these groups, we see almost twice as much sharing during the day on weekdays as at night and on weekends. On weekends we also see slightly greater preferences for sharing during the evening.

About half of our subjects visited 9 or fewer distinct locations throughout the study, and 89% visited 14 or fewer (the max was 27, the min was 3). A subject was considered to have visited a distinct location only if it was visited for at least 15 minutes, and was at least 250 meters from all other locations that the subject visited.

We found that, on average, subjects spent significantly more time at one location than any other (most likely their homes). We also found that the time spent at a location appeared to drop off exponentially for the second, third,

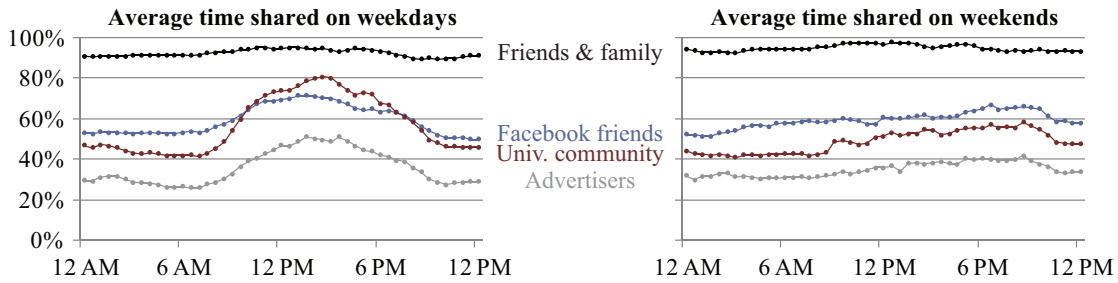

Figure 3: The average percentage of time shared with each group during each thirty-minute interval throughout the day.

fourth and fifth most visited locations. Table 3 shows the average percentage of time a subject spent at his or her three most visited locations, and the average percentage of time that he or she would have shared that location with each of the groups. On average, our subjects were more willing to share their second most visited location than their first. For university community and advertisers they were willing to share it almost twice as often. This suggests that this was most likely a more public location, such as somewhere on or near the university campus.

Location rank (time spent)	Time spent	Time shared w/ group FF	FB	UC	AD
1st	66%	93%	58%	48%	29%
2nd	20%	94%	65%	77%	55%
3rd	6%	90%	61%	62%	41%

Table 3: The average percentage of time a subject spent at his or her three most visited locations, and the average percentage of time he or she would have shared that location with friends and family (FF), Facebook friends (FB), university community (UC), and advertisers (AD).

These results suggest mobility patterns similar to those observed by Gonzalez *et al.*, who found that human trajectories tend to be very patterned, with people visiting a small number of highly frequented places [9]. These results also help explain our later finding that Loc settings only require a few rules to realize most of their benefits.

4.3 Measuring the effects of different privacy-setting types

We will now present analysis quantifying the relative effects of different privacy-setting types, in terms of accuracy and amount of time shared. We consider the results statistically, and under a wide range of assumptions, including varying levels of user burden.

4.3.1 Results regarding policy accuracy

Our first set of results, presented in Figure 4, investigates the accuracy of each of the different privacy-setting types, for each of the groups we asked about. For these results, we hold the cost of mistakenly revealing a location to be fixed at $c = 20$, which is equivalent to assuming that subjects view mistakenly revealing their location as twenty times worse than correctly sharing. We highlight our results for this value of c based on the post-study survey results presented in Table 2, which showed that subjects were significantly concerned with mistakenly revealing

their location to each of the groups other than their close friends and family. Our next set of results will consider varying this cost to account for differences between subjects and groups.

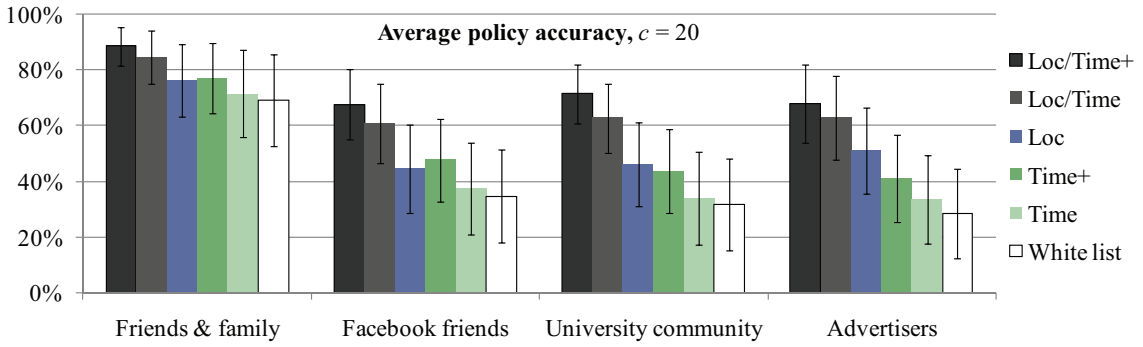

Figure 4: The average accuracy (bars indicate 95% confidence intervals) for each group, under each of the different privacy-setting types. For these results, we hold constant the cost for inappropriately revealing a location at $c = 20$.

Our first observation is that, with $c = 20$, none of the privacy-setting types we consider are able to achieve 100% accuracy for any of the groups. Even the accuracy of the most accurate setting type and group, Loc/Time+ for friends and family, is significantly less than 100%.[14] This demonstrates that a non-trivial subset of our subjects had preferences that alternated between sharing and hiding the same location, at the same time, on different days of the week (most likely due to other contextual factors).

With $c = 20$, the average accuracy of the different privacy-setting types has a wide range across groups, from about 28% (white lists for advertisers) to 88% (Loc/Time+ for friends and family). There is also a moderately large range in accuracy, across groups, for the same simple setting types (e.g., white lists range from 28% to 68%). However, the range across groups is substantially smaller for more complex setting types (e.g., Loc/Time+ settings range from 68% to 88%). This suggests that complex setting types mitigate the importance of a requester's identity in capturing our subjects' preferences.

The range of average accuracies within groups is smaller, but still substantial. For example, within the advertisers group, accuracies range from 68%, for Loc/Time+, to 28%, for white lists. For the Facebook friends and university community groups, we also observe a more than two times increase in accuracy of Loc/Time+ over white lists. The fact that such ranges in accuracy exist within groups further demonstrates that our subjects had diverse privacy preferences that could not all be captured simply by the requester's identity.

For advertisers, the complex setting types (i.e., Loc/Time and Loc/Time+) are significantly more accurate than white lists, Time, and Time+ settings. Loc alone is also significantly better than white lists, and marginally significantly better than Time. The relative importance of location-based rules for this group is consistent with our pre-study survey findings presented in Table 1.

In other groups, we see statistical ties between Loc, Time+, and Time, although Loc tends to be the best of the three on average (primarily due to its effectiveness for advertisers). We also see that the setting types allowing users to distinguish between weekdays and weekends can offer substantial benefits over their simpler counterparts (e.g., for university community Time+ is about 15% more accurate than Time), but these differences are typically not statistically significant.

For university community and Facebook friends, we find that Loc/Time+ is significantly more accurate than any of the other setting types. For university community, we find that Loc/Time is also significantly more accurate than white lists, Time, and Time+, and marginally significantly more accurate than Loc. For Facebook friends the finding

[14] For this we used a one-sample t-test.

is nearly the same, but Time+ is statistically tied with Loc/Time. This demonstrates the importance of weekends in capturing our subjects' preferences about sharing their location with Facebook friends.

All of these results taken together suggest that, with $c = 20$, our subjects could expect significant accuracy improvements from more complex privacy-setting types, and further confirms the hypothesis that the privacy preferences revealed by our study are complex.

Our next set of results, shown in Figure 5, investigates the impact of varying the cost associated with mistakenly revealing a location, for the Facebook friends group. We present these results for Facebook friends only because we believe that this group is of general interest, and results for other groups were qualitatively similar.

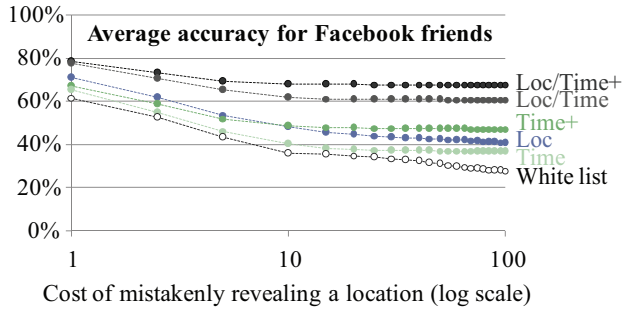

Figure 5: The average accuracy for the Facebook friends group, under each of the different privacy-setting types, while varying the cost associated with mistakenly revealing a location from $c = 1$ to 100.

These results demonstrate that the accuracy benefits of more complex setting types are greatest when information is more sensitive. For example, when $c = 1$, we find that there are no statistically significant differences between any of the setting types. In this case, the difference between the most complex setting type, Loc/Time+, and the simplest, white lists, is only marginally significant. However, the accuracies of simpler setting types drop steeply as the cost of inappropriately revealing one's location increases. For example, the accuracy of white lists drops from 61% at $c = 1$, to almost half of that, or 34%, at $c = 25$, and drops to 28% by the time we reach $c = 100$. Similar patterns are seen with all of the simple setting types, such as Time, Time+, and Loc. This drop is due to the fact that, as this cost goes up, the policies we identify are more restrictive (e.g., by concealing more often). Thus, they provide lower accuracy because they have missed more opportunities to share.

Each of the setting types also reaches a plateau at different values of c. The plateau occurs when the subjects have been forced to hide as much as they can, and only reveal times or locations that are never private. The accuracies of more complex setting types, such as Loc/Time and Loc/Time+, deteriorate far less, far slower, and with plateaus beginning at far lower costs than simple types (e.g., the plateau for Loc/Time+ begins at $c = 10$, whereas white lists continue to lose accuracy throughout the entire range). This demonstrates how more complex setting types can add substantial value for privacy-sensitive users.

4.3.2 Results regarding amount of time shared

We now consider how the policies we identified for different privacy-setting types effect the amount of time our subjects would have shared with each of the groups. Figure 6 shows the average percentage of time that each subject would have shared, under each of the different setting types, with a fixed cost of $c = 20$ for mistakenly revealing a location.

Here we see results similar to those in Figure 4, such that more accurate policies also tend lead to more sharing with each group. For example, for the Facebook friends, university community, and advertiser groups, we see about twice as much sharing with Loc/Time+ settings versus white lists, and in each case this difference is statistically

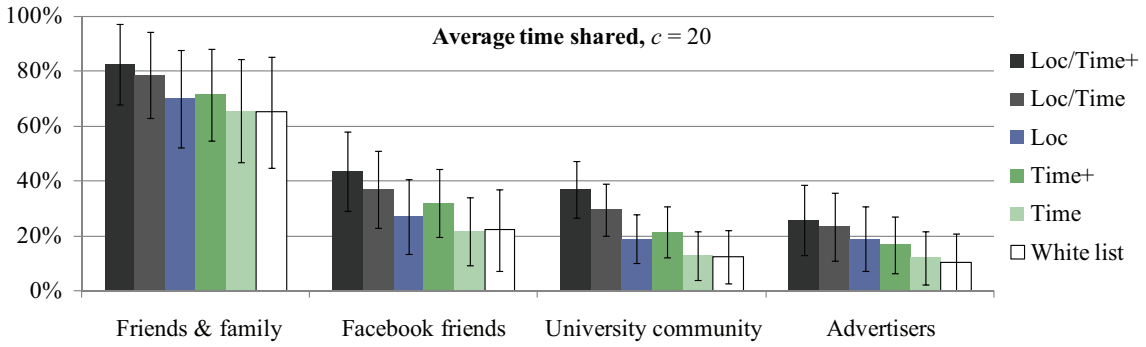

Figure 6: The average percentage of time shared (bars indicate 95% confidence intervals) with each group under each of the different privacy-setting types. For these results, we hold constant the cost for inappropriately revealing a location at $c = 20$.

significant (the difference between Loc/Time and white lists in each case is also marginally significant). It is also interesting to note that Loc and Time+ settings, which are relatively simple, still result in substantial increases in sharing over white lists for the advertiser group (19% and 17% vs. 10%, respectively); however, neither of these differences is statistically significant.

That sharing increases with more accurate setting types is explained by the fact that, when $c = 20$, mistakenly revealing one's location is substantially worse than mistakenly withholding it. This, in turn, leads to policies that tend to err on the safe side and share less.

Our next set of results, presented in Figure 7, considers the effect of varying the cost of mistakenly revealing a location on the amount of time shared under each privacy-setting type. Again, we limit our presentation to the Facebook friends group, since results for other groups were qualitatively the same.

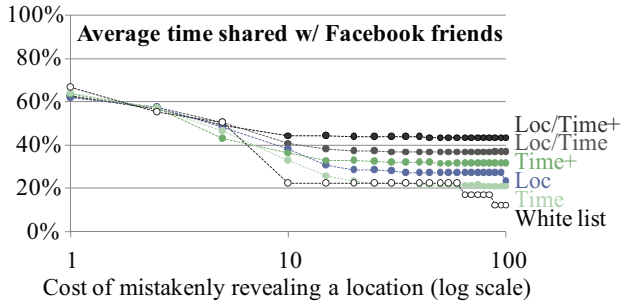

Figure 7: The average percentage of time shared with the Facebook friends group, under each of the different privacy-setting types, while varying the cost associated with mistakenly revealing a location from $c = 1$ to 100.

The findings here are similar to those presented for accuracy in Figure 5, with a few notable differences. We see a general trend from more to less sharing as c increases, with plateaus beginning at around $c = 10$, however the plateaus are far more dramatic and jagged than with accuracy. This is because we only observe effects on sharing when individual rules are made more restrictive, rather than the smooth descent in accuracy that leads to the restriction.

As with accuracy, the decline in sharing with more complex privacy-setting types, such as Loc/Time+ and Loc/Time, is less steep, and slower than that of the simpler types. A higher value for c represents the assumption that users are more concerned about privacy. Thus, this demonstrates how it can actually be in a service's best

interest to offer more complex privacy settings, in order to increase contributions from privacy-sensitive users.

One final take away from this analysis is the magnitude of the increase in sharing with highly privacy-sensitive users, under the most complex setting type, Loc/Time+, versus white lists. For $c = 100$, which corresponds to the assumption that users will make policies that never give out private information, we see a more than three and a half times increase in the average percentage of time shared with the Facebook friends group.

All of these results taken together suggest, somewhat counter-intuitively, that offering richer privacy settings may, in fact, make good business sense, since it will result in privacy-sensitive users sharing more information.

4.3.3 Results under user-burden considerations

In practice, we do not expect users to necessarily specify the most accurate policy matching their preferences, especially under the more complex privacy-setting types, such as Loc/Time+, where user interfaces can be cumbersome. To test the effects of such user-burden considerations on our conclusions, we analyze the effect of limiting the number of rules in policies for each of the setting types.

Our first set of results under user-burden considerations is presented in the four panels of Figure 8, one for each group. It shows the accuracy of each setting type, while varying a limit on the number of rules from one to five or more. This set of results is modeled after a scenario where sharing one's location with all four groups is possible within a single application, and users specify rules that apply to combinations of these groups. We operationalize this by identifying the most accurate policy with a *global rule limit*, rather than a limit that applies to each group individually. For each of the different setting types, we identify policies that equally weight accuracy among the groups.

Unsurprisingly, we find that tighter rule limits generally dampen the accuracy benefits of more complex privacy-setting types. Yet, we see that Loc/Time+ and Loc/Time have substantial benefits, in terms of average global accuracy, with as few as one or two rules. For example, if we consider the global average accuracy across all groups, with only a single rule we already see a marginally significant benefit from Loc/Time+ (51%) over white lists (35%). With two rules, the difference between the accuracy of Loc/Time+ (54%) and white lists is significant, and the difference between the accuracy of Loc/Time (50%) and white lists is marginally significant. This demonstrates how more complex privacy-setting types can be better than simple settings at capturing the preferences of our subjects, while requiring only a small number of rules.

When we examine the effects of a global rule limit on the accuracies within individual groups, rather than the global average accuracy, with two rules we find a significant accuracy improvement for the university community group from Loc/Time+ (52%) over white lists (31%), and a marginally significant difference between those two setting types for advertisers (45% vs. 28%). With three rules, the difference in accuracy between Loc settings (49%) and white lists is significant, and the difference between Loc settings and Time settings (33%) is marginally significant. Interestingly, with three rules, the Loc/Time and Loc/Time+ settings actually perform worse for advertisers than the simpler Loc settings. This is because under the more complex setting types, the three rules are primarily being used to achieve greater accuracy in other groups, whereas the accuracy of Loc tends to plateau with two rules. This plateau can be explained, in part, by the general mobility patterns presented in Table 3, which show that subjects tended to spend about 80% of their time at two distinct locations.

Our final set of results, presented in Figure 9, is modeled after a service where users can share locations with a single group only, such as all of a one's Facebook friends. Here we limit the rules that apply to a group individually, rather than imposing a global limit. We present the results for the Facebook friends group only, but results for other groups were similar.

By comparing the results in Figure 9 to those in the top right panel of Figure 8, we find that with an individual rule limit the accuracy benefits of more complex privacy-setting types are realized with fewer rules. For example, we find that with a single rule the average accuracy benefit of Loc/Time+ (51%) over that of white lists (35%) is

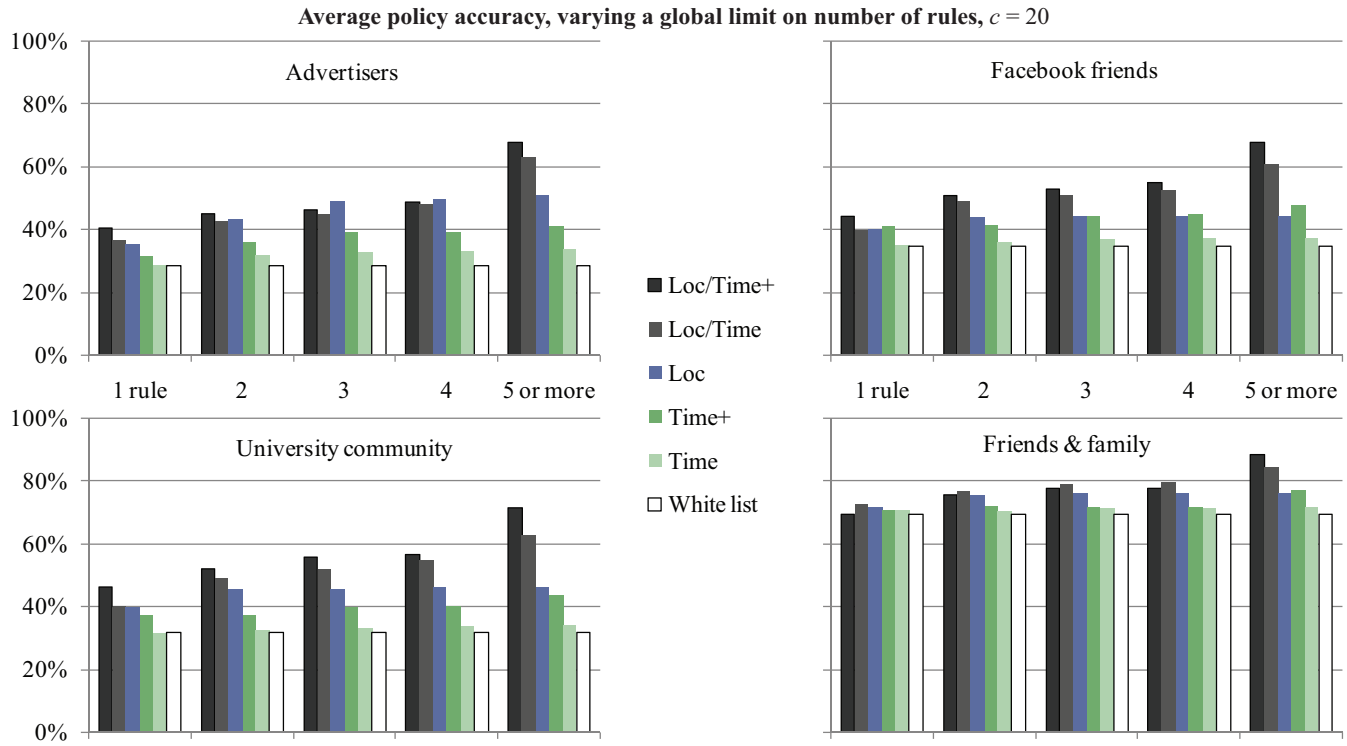

Figure 8: The average accuracy achieved by each of the different privacy-setting types, for each of the different groups, varying a global limit on the number of rules in a policy. We hold constant the cost for inappropriately revealing a location at $c = 20$, and identify policies with the highest possible total accuracy across all groups, while weighting each group equally.

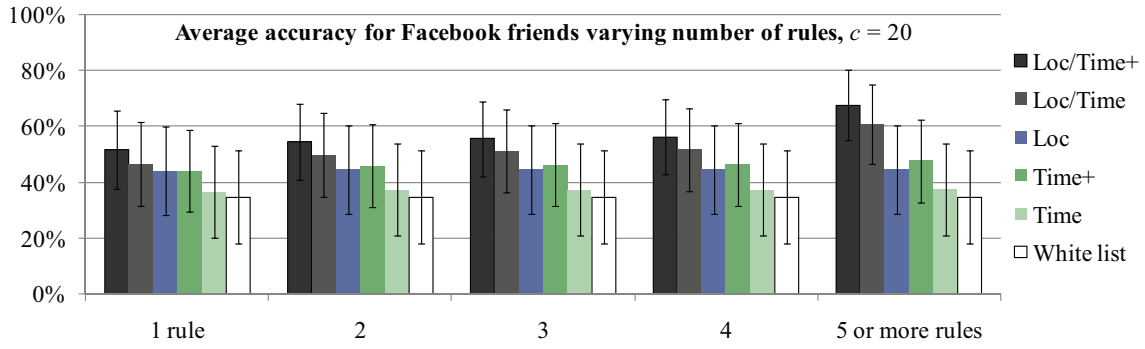

Figure 9: The average accuracy (bars indicate 95% confidence intervals) achieved by each of the different privacy-setting types for the Facebook friends group, while varying a limit on the number of rules in a policy that apply to Facebook friends only. We hold constant the cost for inappropriately revealing a location at $c = 20$

marginally significant, whereas with a global limit it took three rules to reach that level. With a two-rule limit the accuracy benefits of Loc/Time+ (54%) and Loc/Time (50%) over that of white lists are significant and marginally significant, respectively. This demonstrates how complex setting types are likely to be more effective under user-burden considerations in more specialized services.

5 Conclusions and future work

Over the past few years we have seen an explosion in the number and different types of applications that allow individuals to exchange personal information and content that they have created. While there is clearly a demand for users to share this information with each other, they are also demanding greater control over the conditions under which their information is shared.

This paper presented the results from a user study that tracked the locations of 27 subjects over three weeks to collect their stated privacy preferences. Throughout the study, we collected more than 7,500 hours of data. In contrast to some earlier research that identified the requester's identity [7] and user's activity [6] as primarily defining privacy preferences for location sharing, we found that there are a number of other critical dimensions in these preferences, including time of day, day of week, and exact location.

We characterize the complexity of our subjects' preferences by measuring the accuracy of different privacy-setting types. We considered a variety of setting types with differing levels of complexity.

As one might expect, we found that more complex privacy-setting types, such as those that allow users to specify both locations and times at which they are willing to share, were significantly more accurate under a wide variety of assumptions. More surprising was the magnitude of the improvement — in some cases we found an almost three times increase in average accuracy over that of white lists. These findings were also consistent with our pre-study survey, where subjects reported being significantly more comfortable with the prospect of sharing their location using time- and location-based rules.

We also measured the amount of time that our subjects would have shared their location under each of the different privacy-setting types. We found that more complex setting types also generally lead to more sharing. This result, which may at first seem counter intuitive, is due to the fact that users generally tend to err on the safe side, and restrict access with simpler settings. This suggests that offering richer privacy settings may make services more, not less, valuable, by encouraging privacy-sensitive users to share more.

One practical implication of our work is that white lists appear to be very limited in their ability to capture the privacy preferences revealed by our study. This, in combination with the fact that white lists are the only privacy settings offered by most location-sharing applications today (with the notable exception of Locaccino developed by our research group at CMU, which offers all of the privacy-setting types we discussed) [18], suggests that the slow adoption of these services may, in part, be attributed to the simplicity of their privacy settings.

Clearly, as privacy settings become more complex, users may have to spend more time specifying their preferences. To address this, we also examined the impact of the different privacy-setting types under varied assumptions regarding the amount of effort users would be willing to exert while creating their policies. Our findings suggest that, while limiting policies to a small number of rules dampens the accuracy benefits of complex setting types, they generally remain substantially more accurate than white lists.

The user study presented in this paper can be generalized as a methodology for characterizing the tradeoffs between more complex setting types and accuracy in a number of privacy and security domains. At a high level, the methodology involves i) collecting highly detailed preferences from a particular user population, ii) identifying policies for each subject under a variety of different privacy- or security-setting types, and iii) comparing the accuracy of the resulting policies under a variety of assumptions about the sensitivity of the information and tolerance for user burden.

The findings in this paper open several avenues for future work. One avenue involves exploring additional dimensions of privacy preferences. For example, we can study settings that allow users to control the resolution at which location information is provided (e.g., neighborhood, city, or state), or that grant access based on the user's proximity to the requester. We can also investigate the impact of accuracy models that are richer in terms of their tolerance for error. For example, we can use models with costs for mistakenly revealing a location that depend on

the subject, the requester, the time of day, or the location in question.

We examined the impact of a rule limit on the accuracy of more complex privacy-setting types, but we still assumed that users would be able to identify the most accurate possible rules subject to this limit. This opens up another avenue for future work: accounting for additional cognitive limitations, such as bounded rationality [19], to address issues that challenge this assumption. One potential method for accomplishing this would be to study the behavior of real users of a location-sharing application that offers all of the different privacy-setting types discussed in this paper, such as Locaccino. We could then compare actual user behavior to the predictions of our models, and better characterize the difference between what is predicted by our analysis and what users will actually do in practice.

6 Acknowledgments

This work has been supported by NSF grants CNS-0627513, CNS-0905562, and DGE-0903659. Additional support has been provided by Nokia, France Telecom, Google, the CMU/Microsoft Center for Computational Thinking, ARO research grant DAAD19-02-1-0389 to Carnegie Mellon University's CyLab, and the CMU/Portugal Information and Communication Technologies Institute. The authors would also like to thank Paul Hankes-Drielsma, Janice Tsai, Tuomas Sandholm, Lucian Cesca, Jialiu Lin, Tony Poor, Eran Toch, and Kami Vaniea for their assistance with our study.

References

[1] L. Barkhuus, B. Brown, M. Bell, M. Hall, S. Sherwood, and M. Chalmers. From awareness to repartee: Sharing location within social groups. In *Conference on Human Factors in Computing Systems (CHI)*, 2008.

[2] L. Barkhuus and A. Dey. Location-based services for mobile telephony: A study of users' privacy concerns. In *International Conference on Human-Computer Interaction (INTERACT)*, 2003.

[3] M. Benisch, N. Sadeh, and T. Sandholm. A theory of expressiveness in mechanisms. In *National Conference on Artificial Intelligence (AAAI)*, 2008.

[4] M. Benisch, N. Sadeh, and T. Sandholm. Methodology for designing reasonably expressive mechanisms with application to ad auctions. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2009.

[5] T. Burghardt, E. Buchmann, J. Müller, and K. Böhm. Understanding user preferences and awareness: Privacy mechanisms in location-based services. In *OnTheMove Conferences (OTM)*, 2009.

[6] K. Connelly, A. Khalil, and Y. Liu. Do I do what I say?: Observed versus stated privacy preferences. In *INTERACT*, 2007.

[7] S. Consolovo, I. Smith, T. Matthews, A. LaMarca, J. Tabert, and P. Powledge. Location disclosure to social relations: Why, when, & what people want to share. In *CHI*, 2005.

[8] J. Cornwell, I. Fette, G. Hsieh, M. Prabaker, J. Rao, K. Tang, K. Vaniea, L. Bauer, L. Cranor, J. Hong, B. McLaren, M. Reiter, and N. Sadeh. User-controllable security and privacy for pervasive computing. In *Workshop on Mobile Computing Systems and Applications*, 2007.

[9] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[10] K. Group. BIA's The Kelsey Group Forecasts U.S. Mobile Local Search Advertising Revenues to reach $1.3B in 2013, February 2009. http://www.kelseygroup.com/press/.

[11] J. Hightower, A. LaMarca, and I. E. Smith. Practical lessons from Place Lab. *IEEE Pervasive Computing*, 5(3):32–39, 2006.

[12] S. Huang, F. Proulx, and C. Ratti. iFIND: a Peer-to-Peer application for real-time location monitoring on the MIT campus. In *International Conference on Computers in Urban Planning and Urban Management (CUPUM)*, 2007.

[13] G. Iachello, I. Smith, S. Consolovo, G. Abowd, J. Hughes, J. Howard, F. Potter, J. Scott, T. Sohn, J. Hightower, and A. LaMarca. Control, deception, and communication: Evaluating the deployment of a location-enhanced messaging service. In *UbiComp*, 2005.

[14] S. Lederer, J. Mankoff, and A. K. Dey. Who wants to know what when? privacy preference determinants in ubiquitous computing. In *CHI*, 2003.

[15] M. Mazurek, J. Arsenault, J. Bresee, N. Gupta, I. Ion, C. Johns, D. Lee, Y. Liang, J. Olsen, B. Salmon, R. Shay, K. Vaniea, L. Bauer, L. Cranor, G. Ganger, and M. Reiter. Access control for home data sharing: Attitudes, needs and practices. In *CHI*, 2010.

[16] S. Patil and J. Lai. Who gets to know what when: Configuring privacy permissions in an awareness application. In *CHI*, 2005.

[17] N. Sadeh, F. Gandon, and O. B. Kwon. Ambient intelligence: The MyCampus experience. In T. Vasilakos and W. Pedrycz, editors, *Ambient Intelligence and Pervasive Computing*. ArTech House, 2006.

[18] N. Sadeh, J. Hong, L. Cranor, I. Fette, P. Kelley, M. Prabaker, and J. Rao. Understanding and capturing people's privacy policies in a mobile social networking application. *The Journal of Personal and Ubiquitous Computing*, 13(6):401–412, 2009.

[19] H. A. Simon. *Models of Man*. John Wiley & Sons, 1957.

[20] I. Smith, S. Consolovo, A. LaMarca, J. Hightower, J. Scott, T. Sohn, J. Hughes, G. Iachello, and G. Abowd. Social disclosure of place: From location technology to communication practices. In *Lecture Notes in Computer Science : Pervasive Computing*, pages 134–151, 2005.

[21] J. Tsai, P. Kelley, L. Cranor, and N. Sadeh. Location-sharing technologies: Privacy risks and controls. In *Research Conference on Communication, Information and Internet Policy (TPRC)*, 2009.

[22] J. Tsai, P. Kelley, P. H. Drielsma, L. F. Cranor, J. Hong, and N. Sadeh. Who's viewed you? the impact of feedback in a mobile-location system. In *CHI*, 2009.

[23] Y. Wang, J. Lin, M. Annavaram, Q. A. Jacobson, J. Hong, B. Krishnamachari, and N. Sadeh. A framework of energy efficient mobile sensing for automatic user state recognition. In *MobiSys*, 2009.

[24] R. Want, V. Falcão, and J. Gibbons. The active badge location system. *ACM Transactions on Information Systems*, 10:91–102, 1992.